

Whole-Genome Sequences of 94 Environmental Isolates of *Bacillus cereus Sensu Lato*

Géraldine A. Van der Auwera,^{a,b} Michael Feldgarden,^b Roberto Kolter,^c Jacques Mahillon^a

Laboratory of Food & Environmental Microbiology, Université Catholique de Louvain, Louvain-la-Neuve, Belgium^a; Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA^b; Kolter Laboratory, Harvard Medical School, Boston, Massachusetts, USA^c

Bacillus cereus sensu lato is a species complex that includes the anthrax pathogen *Bacillus anthracis* and other bacterial species of medical, industrial, and ecological importance. Their phenotypes of interest are typically linked to large plasmids that are closely related to the anthrax plasmids pXO1 and pXO2. Here, we present the draft genome sequences of 94 isolates of *B. cereus sensu lato*, which were chosen for their plasmid content and environmental origins.

Received 5 May 2013 Accepted 28 August 2013 Published 3 October 2013

Citation Van der Auwera GA, Feldgarden M, Kolter R, Mahillon J. 2013. Whole-genome sequences of 94 environmental isolates of *Bacillus cereus sensu lato*. *Genome Announc.* 1(5):e00380-13. doi:10.1128/genomeA.00380-13.

Copyright © 2013 Van der Auwera et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](http://creativecommons.org/licenses/by/3.0/).

Address correspondence to Géraldine A. Van der Auwera, vdauwera@broadinstitute.org.

Bacillus cereus sensu lato is a species complex that groups six classically described species of ubiquitous Gram-positive spore-forming bacteria, including the eponymous *B. cereus sensu stricto*, the entomopathogen *Bacillus thuringiensis*, the rhizoid-looking *Bacillus mycoides* and *Bacillus pseudomycoides*, and *Bacillus anthracis*, the etiologic agent of anthrax.

The members of this group were originally distinguished on the basis of their phenotypic differences, but over the past decade, advances in the understanding of the phylogenomics of this group have largely invalidated this classification. Instead, the members of the *B. cereus sensu lato* group are more appropriately viewed as forming one single species from which different ecotypes and pathotypes emerge in a dynamic fashion, leading in some cases to the formation of clonal complexes with specific phenotypes (1–6).

Many phenotypic properties that are specific to these ecotypes and pathotypes are directly related to the presence or absence of large plasmids that carry genes associated with those phenotypes. In the case of *B. anthracis*, the virulence plasmids pXO1 (192 kb) and pXO2 (96 kb) carry the anthrax toxin and capsule genes, respectively, as well as the associated regulatory elements (7). Furthermore, the large plasmids found in a number of previously sequenced *B. cereus sensu lato* strains of medical or industrial interest were observed to share a backbone with either the pXO1 or pXO2 anthrax plasmids. For example, in strains of *B. cereus sensu stricto* that are responsible for the *B. cereus*-associated emetic food poisoning syndrome, the genes encoding the emetic toxin cereulide are carried by a large plasmid that shares a common genetic backbone with the pXO1 anthrax plasmid (8, 9).

We have shown previously that select sequences of the shared pXO1 and pXO2 backbones can be found widely in environmental isolates of *B. cereus sensu lato* (10). We postulated that these are found in plasmids that are genetically related to the pXO1 and pXO2 plasmids (hence called pXO1-like and pXO2-like, respec-

tively) and may play an important role in the ecotypic and pathotypic differentiation of *B. cereus sensu lato* organisms.

In order to gain deeper insight into the ecological distribution and genomic diversity of the pXO1-like and pXO2-like plasmids, we sequenced a panel of 94 isolates of *B. cereus sensu lato* organisms containing a variety of plasmids and having diverse environmental origins. This adds to the ~60 whole or draft genomes of *B. cereus* of various origins already available in GenBank.

De novo assemblies were generated from Illumina 101-base paired-end reads generated with two libraries, one from 180-bp fragments and one from 3-kb jumping libraries. The assemblies were constructed using AllPaths-LG (11). The protein-coding genes were predicted with Prodigal (12) and filtered to remove genes with ≥70% overlap to the tRNAs or rRNAs. The tRNAs were identified by tRNAscan-SE (13). The rRNA genes were predicted using RNAmmer (14). The gene product names were assigned based on top BLAST hits against the Swiss-Prot protein database (≥70% identity and ≥70% query coverage) and a protein family profile search against the TIGRfam HMMER equivalents.

Nucleotide sequence accession numbers. All 94 draft genome sequences have been deposited at GenBank under the accession no. reported in Table 1 (10, 15–22).

ACKNOWLEDGMENTS

This project was funded in part with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under contract no. HHSN272200900018C. G.A.V.D.A. was supported by postdoctoral fellowships from the Belgian American Educational Foundation (BAEF) and from the Fonds National pour la Recherche Scientifique (FNRS) of Belgium.

G.A.V.D.A. and J.M. thank the many collaborators at UCL Louvain-la-Neuve, Harvard Medical School, and Huazhong Agricultural University who contributed environmental isolates to this study.

TABLE 1 Strain characteristics

Sample source	Strain name	GenBank accession no.	Predicted plasmid(s) ^d	Reference
Soil, Greenland	VD048	AHEU01000000	pXO1	10
	VD078	AHEV01000000	pXO1	10
	VD045	AHET01000000	pXO2	10
	VDM022	AHFP01000000	pXO2	10
	VDM021	AHFU01000000	Neither	10
	VDM019	AHFO01000000	Neither	10
Soil, Spain	VD014	AHER01000000	pXO1	10
	VDM006	AHFT01000000	pXO2	10
	VDM034	AHFQ01000000	Neither	10
Soil, Scotland	VD142	AHCL01000000	pXO2	10
	VD148	AHFF01000000	pXO2	10
	VDM062	AHFS01000000	pXO2	10
	VD136	AHFC01000000	Neither	10
	VD140	AHFD01000000	Neither	10
	VD146	AHFE01000000	Neither	10
Water, Scotland	VD200	AHFM01000000	pXO1	10
	VD214	AHFN01000000	Neither	10
Soil, Martinique	VD133	AHFB01000000	pXO1	10
	VD131	AHFA01000000	Neither	10
Soil, Guadeloupe	VD107	AHEX01000000	pXO2	10
	VD115	AHEY01000000	pXO2	10
	VD102	AHEW01000000	Neither	10
	VD118	AHEZ01000000	Neither	10
Soil, Abu Dhabi, UAE	VD156	AHFH01000000	pXO1	10
	VD154	AHFG01000000	Neither	10
Soil, Dubai, UAE	VD169	AHFJ01000000	pXO1	10
	VD196	AHFL01000000	pXO1	10
	VD166	AHFI01000000	Neither	10
	VD184	AHFK01000000	Neither	10
Water, Belgium (small pond site)	VD021	AHES01000000	pXO2	10
	VD022	AHCK01000000	pXO1, pXO2	10
	VDM053	AHFR01000000	Neither	10
Soil, Belgium (site A)	HuA2-1	AHDV01000000	pXO1	10
	HuA2-3	AHDW01000000	Neither	10
	HuA2-4	AHDX01000000	pXO2	10
	HuA2-9	AHDY01000000	Neither	10
	HuA3-9	AHDZ01000000	pXO1	10
	HuA4-10	AHEA01000000	pXO2	10
Soil, Belgium (site B)	HuB1-1	AHEB01000000	Neither	10
	HuB2-9	AHED01000000	pXO2	10
	HuB4-4	AHEF01000000	Neither	10
	HuB4-10	AHEE01000000	pXO2	10
	HuB5-5	AHEG01000000	pXO2	10
	HuB13-1	AHEC01000000	Neither	10
Soil, Massachusetts (Boston site AG)	BAG1O-1	AHCN01000000	Neither	This study
	BAG1O-2	AHCO01000000	Neither	This study
	BAG1O-3	AHCP01000000	Neither	This study
	BAG1X1-1	AHCQ01000000	pXO1	This study
	BAG1X1-2	AHCR01000000	pXO1	This study
	BAG1X1-3	AHCS01000000	pXO1	This study
	BAG1X2-1	AHCT01000000	pXO1, pXO2	This study
	BAG1X2-2	AHCU01000000	pXO2	This study
	BAG1X2-3	AHCV01000000	pXO2	This study

(Continued on following page)

Table 1 (Continued)

Sample source	Strain name	GenBank accession no.	Predicted plasmid(s) ^d	Reference
	BAG2O-1	AHCW01000000	Neither	This study
	BAG2O-2	AHCX01000000	Neither	This study
	BAG2O-3	AHCY01000000	Neither	This study
	BAG2X1-1	AHCZ01000000	pXO1	This study
	BAG2X1-2	AHDA01000000	pXO1	This study
	BAG2X1-3	AHDB01000000	pXO1	This study
	BAG3O-1	AHFV01000000	Neither	This study
	BAG3O-2	AHDC01000000	Neither	This study
	BAG3X2-1	AHDD01000000	pXO2	This study
	BAG3X2-2	AHDE01000000	pXO2	This study
	BAG4O-1	AHDF01000000	Neither	This study
	BAG4X2-1	AHDH01000000	pXO2	This study
	BAG4X12-1	AHDG01000000	pXO1, pXO2	This study
	BAG5O-1	AHDI01000000	Neither	This study
	BAG5X1-1	AHDJ01000000	pXO1	This study
	BAG5X2-1	AHDL01000000	pXO2	This study
	BAG5X12-1	AHDK01000000	pXO1, pXO2	This study
Soil, Massachusetts (Boston site ES)	BAG6O-1	AHDM01000000	Neither	This study
	BAG6O-2	AHDN01000000	Neither	This study
	BAG6X1-1	AHDO01000000	pXO1	This study
	BAG6X1-2	AHDP01000000	pXO1	This study
Food	AND1407	AHCM01000000	pXO1	15
	K-5975c	AHEL01000000	pXO1	19
	TIAC219	AHCJ01000000	pXO1	20
	Schrouff	AHCI01000000	pXO1, pXO2	Mahillon et al., unpublished
	ISP3191	AHEK01000000	pXO1, pXO2	Dierick et al., unpublished
	ISP2954	AHEJ01000000	pXO1	Dierick et al., unpublished
Soil, China	B5-2	AHFW01000000	pXO1, pXO2	Sun et al., unpublished
Mammals, Poland	IS075	AHCH01000000	pXO1, pXO2	21
	IS195	AHEH01000000	pXO1, pXO2	21
	IS845/00	AHEI01000000	pXO1	21
Insect ^a	HD73	AH DU01000000	pXO2	22
Various origins ^b	MC67	AHEN01000000	Other	17
	MC118	AHEM01000000	Other	17
	MSX-A1	AHEO01000000	Other	16
	MSX-A12	AHEP01000000	Other	16
	MSX-D12	AHEQ01000000	Other	16
	BMG1.7	AHDQ01000000	Other	18
Various origins ^c	CER057	AHDS01000000		15
	CER074	AHDT01000000		15
	BtB2-4	AHDR01000000		15

^a Reference strain for the pXO2-like plasmid pAW63.^b Strains predicted to carry other plasmids of interest.^c Strains predicted to carry plasmid fragments on the chromosome.^d Neither, neither pXO1 no pXO2; other, plasmids other than pXO1 or pXO2.

REFERENCES

- Helgason E, Okstad OA, Caugant DA, Johansen HA, Fouet A, Mock M, Hegna I, Kolstø AB. 2000. *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*—one species on the basis of genetic evidence. 66:2627–2630.
- Priest FG, Barker M, Baillie LWJ, Holmes EC, Maiden MCJ. 2004. Population structure and evolution of the *Bacillus cereus* group. J. Bacteriol. 186:7959–7970.
- Rasko DA, Altherr MR, Han CS, Ravel J. 2005. Genomics of the *Bacillus cereus* group of organisms. FEMS Microbiol. Rev. 29:303–329.
- Tourasse NJ, Helgason E, Okstad OA, Hegna IK, Kolstø AB. 2006. The *Bacillus cereus* group: novel aspects of population structure and genome dynamics. J. Appl. Microbiol. 101:579–593.
- Maughan H, Van der Auwera G. 2011. *Bacillus* taxonomy in the genomic era finds phenotypes to be essential though often misleading. Infect. Genet. Evol. 11:789–797.
- Zwick ME, Joseph SJ, Didelot X, Chen PE, Bishop-Lilly KA, Stewart AC, Willner K, Nolan N, Lentz S, Thomason MK, Sozhamannan S, Mateczun AJ, Du L, Read TD. 2012. Genomic characterization of the *Bacillus cereus sensu lato* species: backdrop to the evolution of *Bacillus anthracis*. Genome Res. 22:1512–1524.
- Koehler TM. 2009. *Bacillus anthracis* physiology and genetics. Mol. Aspects Med. 30:386–396.
- Hoton FM, Andrup L, Swiecicka I, Mahillon J. 2005. The cereulide genetic determinants of emetic *Bacillus cereus* are plasmid-borne. 151: 2121–2124.

9. Ehling-Schulz M, Fricker M, Grallert H, Rieck P, Wagner M, Scherer S. 2006. Cereulide synthetase gene cluster from emetic *Bacillus cereus*: structure and location on a mega virulence plasmid related to *Bacillus anthracis* toxin plasmid pXO1. 6:20.
10. Hu X, Van der Auwera G, Timmerly S, Zhu L, Mahillon J. 2009. Distribution, diversity, and potential mobility of extrachromosomal elements related to the *Bacillus anthracis* pXO1 and pXO2 virulence plasmids. *Appl. Environ. Microbiol.* 75:3016–3028.
11. Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. 108:1513–1518.
12. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi:10.1186/1471-2105-11-119.
13. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25: 955–964.
14. Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35:3100–3108.
15. Hoton FM, Fornelos N, N'Guessan E, Hu X, Swiecicka I, Dierick K, Jääskeläinen E, Salkinoja-Salonen M, Mahillon J. 2009. Family portrait of *Bacillus cereus* and *Bacillus weihenstephanensis* cereulide-producing strains. *Environ. Microbiol. Rep* 1:177–183.
16. Timmerly S, Hu X, Mahillon J. 2011. Characterization of bacilli isolated from the confined environments of the Antarctic Concordia station and the International Space Station. *Astrobiology* 11:323–334.
17. Thorsen L, Hansen BM, Nielsen KF, Hendriksen NB, Phipps RK, Budde BB. 2006. Characterization of emetic *Bacillus weihenstephanensis*, a new cereulide-producing bacterium. *Appl. Environ. Microbiol.* 72: 5118–5121.
18. Cherif A, Ouzari H, Daffonchio D, Cherif H, Ben Slama K, Hassen A, Jaoua S, Boudabous A. 2001. Thuricin 7: a novel bacteriocin produced by *Bacillus thuringiensis* BMG1.7, a new strain isolated from soil. *Lett. Appl. Microbiol.* 32:243–247.
19. Dierick K, Van Coillie E, Swiecicka I, Meyfroidt G, Devlieger H, Meulemans A, Hoedemaekers G, Fourie L, Heyndrickx M, Mahillon J. 2005. Fatal family outbreak of *Bacillus cereus*-associated food poisoning. *J. Clin. Microbiol.* 43:4277–4279.
20. Naranjo M, Denayer S, Botteldoorn N, Delbrassinne L, Veys J, Wae-genaere J, Sirtaine N, Driesen RB, Sipido KR, Mahillon J, Dierick K. 2011. Sudden death of a young adult associated with *Bacillus cereus* food poisoning. *J. Clin. Microbiol.* 49:4379–4381.
21. Swiecicka I, De Vos P. 2003. Properties of *Bacillus thuringiensis* isolated from bank voles. *J. Appl. Microbiol.* 94:60–64.
22. Wilcks A, Jayaswal N, Lereclus D, Andrup L. 1998. Characterization of plasmid pAW63, a second self-transmissible plasmid in *Bacillus thuringiensis* subsp. *kurstaki* HD73. *Microbiology* 144(Pt 5):1263–1270.